



Porter, R. J., Tadic, V., & Achim, A. (2016). Better than ℓ_0 recovery via blind identification. In *2015 IEEE Global Conference on Signal and Information Processing, GlobalSIP 2015: Proceedings of a meeting held 14-16 December 2015, in Orlando, FL, USA* (pp. 1280-1284). [7418404] Institute of Electrical and Electronics Engineers (IEEE). <https://doi.org/10.1109/GlobalSIP.2015.7418404>

Peer reviewed version

Link to published version (if available):
[10.1109/GlobalSIP.2015.7418404](https://doi.org/10.1109/GlobalSIP.2015.7418404)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via IEEE at <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=7418404>. Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

Better Than ℓ_0 Recovery via Blind Identification

Richard Porter^{1,a}, Vladislav Tadic^{2,b}, Alin Achim^{3,a}

¹R.Porter@bristol.ac.uk, ²v.b.tadic@bristol.ac.uk, ³eeama@bristol.ac.uk

^aUniversity of Bristol, Department of Electrical & Electronic Engineering, Merchant Venturers Building, Woodland Road, Bristol, BS8 1UB

^bUniversity of Bristol, Department of Mathematics, University Walk, Bristol, BS8 1TW

Abstract—In this work, we propose a novel approach to multiple measurement vector (MMV) compressed sensing. We show that by exploiting the statistical properties of the sources, we can do better than previously derived lower bounds in this context. We show that in the MMV case, we can identify the active sources with fewer sensors than sources. We first develop a general framework for recovering the sparsity profile of the sources by combining ideas from compressed sensing with blind identification methods. We do this by comparing the large known sensing matrix to the smaller matrix estimated by a blind identification method. Finally, we demonstrate the performance of this technique with a variety of data and blind identification methods, and show that under certain assumptions, it is possible to identify the active sources with only 2 sensors, regardless of the number of sources.

Index Terms—Compressed Sensing, Blind identification

I. INTRODUCTION

Work on compressed sensing began with the work of Candès, Romberg, Tao [1], and Donoho [2] on the single measurement vector (SMV) model, and it was later demonstrated that improved results could be obtained by employing a multiple measurement vector (MMV) model [3]. The latter constitutes the focus of this work.

Compressed sensing techniques are typically based on computationally efficient methods for approximating the minimisation of the ℓ_0 pseudonorm. However, as we show in this work ℓ_0 minimisation is not necessarily optimal.

The related problem of Blind Source Separation (BSS) was first initiated through work on motion decoding in vertebrates [4]. It has found uses in medical imaging [5], signal processing [6] and electrocardiogram [7]. A number of algorithms [8], [9] have been developed that are capable of recovering the sources up to permutation and scaling (and without further information, this is the best that can be done). The problem of underdetermined BSS is much more difficult, and typically it is only possible to accurately recover the mixing model, which corresponds to the problem of blind identification.

Blind identification is very closely related to BSS. Although it has been known for quite some time that underdetermined blind identification methods are possible [10],

it has not been until much more recently that practical methods have been demonstrated for the general case [11], [12], although there were some methods proposed for specific cases and sparsity-based approaches [13], [14], [15].

It has been previously shown that the performance of compressive sensing algorithms can be improved by using partial information about the support [16], [17], and our proposed method provides a way of estimating this in the MMV case. To the best of our knowledge, there are no existing algorithms capable of estimating the sparsity profile in the MMV case when there are at most as many sensors as sources.

The rest of this work is organised as follows. In Section II we introduce the mixing model we use. In Section III we introduce a general framework for recovering the sparsity profile. In Section IV we introduce several preexisting blind identification algorithms, and in Section V we demonstrate the effectiveness of applying these methods to suitable data. Finally, we conclude our work in Section VI.

II. BACKGROUND

The SMV model is given by

$$x = As + v \quad (1)$$

Here, $s \in \mathbb{R}^N$ is the source vector, $v \in \mathbb{R}^M$ is the noise vector which is typically assumed to follow a Gaussian distribution, $x \in \mathbb{R}^M$ is the vector of observations, and $A \in \mathbb{R}^{M \times N}$ is the sensing matrix.

It has been shown that under certain assumptions on A [18] if s is k -sparse in the sense that at most k of its entries are non-zero, $2k$ measurements are sufficient to recover s exactly. However, it has also been shown that for practical recovery algorithms we need $\mathcal{O}(k \log(\frac{N}{k}))$ measurements.

On the other hand, the following is a standard MMV CS problem:

$$X = AS + V \quad (2)$$

Here, $S \in \mathbb{R}^{N \times L}$ is the source matrix, $V \in \mathbb{R}^{M \times L}$ is the noise matrix which is typically assumed to follow a Gaussian distribution, $x \in \mathbb{R}^{M \times L}$ is the matrix of observations, and $A \in \mathbb{R}^{M \times N}$ is the sensing matrix. In the MMV case, we call S k -sparse if k rows of S are non-zero. Similarly, when

referring to ℓ_0 pseudonorm minimisation in this context, we are referring to minimising the number of non-zero rows.

It has been shown that in the MMV case, if the columns of S share the same sparsity profile and are k -sparse (i.e. at most k non-zero components), we can relax the bounds from the SMV case and $M = k + 1$ measurements are sufficient to guarantee exact recovery [3]. We can also see that this is the best that can be done by minimising the number of non-zero rows. If we assume S is k -sparse and that we have only k measurements (i.e. $N = k$), then we can pick any arbitrary set of k rows of S , and then take \tilde{A} to be the matrix we obtain by taking the corresponding rows of A . We now take $\hat{\tilde{S}} = \tilde{A}^{-1}X$, and then construct an estimate of S , \hat{S} by mapping the rows of $\hat{\tilde{S}}$ back to the set of k rows we chose. By construction, we have that $A\hat{S} = \hat{A}\hat{\tilde{S}} = X$, hence \hat{S} is a valid k -sparse solution. However, since the selection of the set of k rows was arbitrary, in general $\hat{S} \neq S$ and therefore k measurements are insufficient to recover S if our approach is to minimise the number of non-zero rows.

By notation, A_i denotes the i^{th} row of the matrix A and A_j the j^{th} column of A .

III. METHOD

Let us note that equation 2 can be rewritten as

$$X = \tilde{A}\tilde{S} + V \quad (3)$$

Here, \tilde{S} is S after all the zero rows have been removed, and \tilde{A} is A after all the columns of A corresponding to the zero rows of S are removed. \tilde{A} is unknown, but we know each column of \tilde{A} is equal to a column of A .

It has been known for some time that under some weak assumptions blind identification algorithms can recover the mixing matrix up to permutation and scaling if $M = 2$ [10], no matter how large N is, with the accuracy of recovery improving as $L \rightarrow \infty$. However, we can use the knowledge of A to resolve the permutation and scaling ambiguities, and therefore accurately recover the sparsity profile.

To solve the scaling ambiguity, we rescale the columns of A (and therefore \tilde{A}) by dividing each column by its first entry and then rescaling to unit norm. This works even if A is complex.

The algorithm for identifying the sources that are active (non-zero) as follows:

- 1) Choose the Blind Identification algorithm most appropriate for the nature of your data, and use it to recover an estimate of \tilde{A} . Call this estimate $\hat{\tilde{A}}$.
- 2) Rescale $\hat{\tilde{A}}$ and A by dividing each column of each matrix by its first entry and scaling to unit norm.
- 3) For each column of $\hat{\tilde{A}}$, $\hat{\tilde{A}}_j$ find the j that minimises (4) and say that the j^{th} source is active.

$$\left\| \hat{\tilde{A}}_j - A_j \right\|_2 \quad (4)$$

Where A_i denotes the i^{th} row of the matrix A and A_j the j^{th} column of A .

Assuming that the entries of A are drawn from a continuous probability distribution, with probability 1:

$$\min_{i,j,i \neq j} (\|A_i - A_j\|_2) = \varepsilon > 0 \quad (5)$$

Alternatively, this condition is also satisfied if $\text{spark}(A) > 2$. Note that both the Restricted Isometry Property (RIP) and the weaker Null Space Property (NSP) commonly found in compressed sensing also imply this condition. Hence, if the recovered column is within $\frac{\varepsilon}{2}$ of the real column, it will be matched to the correct column of A . Therefore, assuming that the blind identification algorithm used converges in the following sense

$$\mathbb{P}(\max_i (\min_j (\|\hat{\tilde{A}}_i - \hat{A}_j\|_2)) \geq \frac{\varepsilon}{2}) \rightarrow 0 \text{ as } L \rightarrow \infty \quad (6)$$

We have that the set of active sources is correctly identified with probability tending to 1 as $L \rightarrow \infty$

IV. BLIND IDENTIFICATION ALGORITHMS

In this section, we provide a brief overview of the blind identification algorithms that we will be using to test our approach.

A. FastICA

The case $M = k$ corresponds to the determined BSS case, and therefore we have the choice of a vast variety of options. We have chosen to use the popular FastICA algorithm [8].

FastICA is an Independent Component Analysis (ICA) algorithm, requiring the sources to be non-Gaussian and independent. Given $X = AS + V$, ICA algorithms work by finding an unmixing matrix W such that the rows of WX are statistically independent. The FastICA algorithm works by minimising the mutual information through approximation of the negentropy.

Although the number of sensors is reduced by only one from the case of ℓ_0 pseudonorm minimisation (we need k as opposed to $k + 1$ sensors), this approach is still of significant interest because it is still a determined problem, and therefore recovering the sparsity profile is equivalent to recovering the sources. The intuitive explanation is that we are effectively imposing the dual conditions that S must have k non-zero rows, and that these rows must be independent.

B. SOBIUM

For our blind identification algorithm, we used the Second Order Blind Identification of Underdetermined Mixtures (SOBIUM) algorithm [12]. This can identify the sources so long as $2k(k - 1) \leq M^2(M - 1)^2$, i.e. the number of sensors required is proportional to the square root of the number of active sources. The SOBIUM algorithm requires that the sources be temporally correlated.

The SOBIUM algorithm works by calculating the covariance matrices given by

$$C_p = \mathbb{E}\{x_t x_{t+\tau_p}^H\}, \quad p = 1, \dots, P \quad (7)$$

These matrices are then stacked in a tensor in $\mathbb{C}^{M \times M \times P}$, and the canonical rank-1 decomposition in k terms is calculated, using the alternating least squares approach. Although [12] states that one of the delays τ_p can be equal to zero, this is in fact only true for high SNR, as can be clearly seen in Figure 3.

C. ALESCAF

Finally, we examine the case with the absolute minimum number of sensors, $M = 2$. For this case, we use the ALESCAF algorithm [11].

In addition to the assumptions already made, we assume that the sources are non-Gaussian and that their second characteristic functions are finite, and have finite non-vanishing derivatives up to order 3 in a region around the origin.

Blind identification using the second characteristic function was first proposed in [19] as an algebraic method for $M = 2$ sensors and an arbitrary number of sources, and was later extended for an arbitrary number of sensors in [11], where the more robust tensor decomposition approach was also proposed. The ALESCAF algorithm uses a similar tensor factorisation approach to the SOBIUM approach, but instead of estimating the covariance matrices, we estimate the derivatives of the second characteristic function. When using only third order derivatives, ALESCAF requires that $k \leq 3M - 3$. However with higher order derivatives this can be relaxed. The ALGECF algorithm which is also described in [11] has no theoretical limit on k , but does require computing derivatives up to order k which may be problematic.

V. RESULTS

For the sake of simplicity, we will assume that the number of non-zero of S , denoted by k is known.

The success rate is defined as the number of truly active sources that were marked as active divided by the total number of truly active sources.

A. FastICA

For testing our approach with FastICA, we used the FastICA MATLAB software package [20]. We used the tanh nonlinearity option.

In our experiments, S is a 6×5000 matrix, of which 3 of the rows are non-zero, each corresponding to the real part of distinct quadrature phase-shift keying (QPSK) signals.

The results can be seen in Figures 1 and 2. Figure 1 shows how the success rate changes as the noise varies, and Figure 2 shows how the success rate varies as the number of samples changes with a signal-to-noise ratio (SNR) of 50 dB.

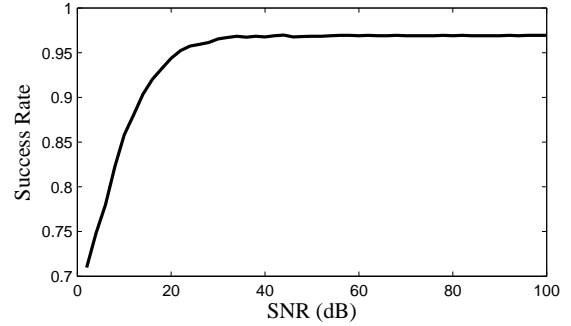


Figure 1. Success Rate of recovering sparsity profile with FastICA as noise varies

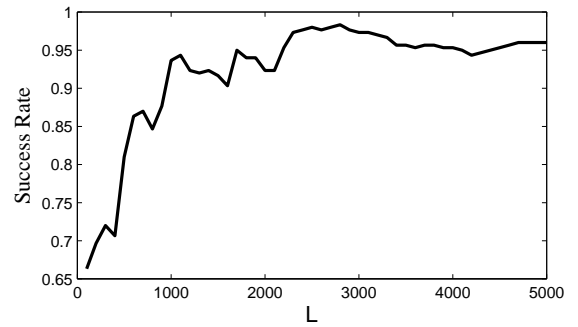


Figure 2. Success rate of recovering sparsity profile with FastICA as L varies

B. SOBIUM

For testing our approach with SOBIUM, we implemented the SOBIUM algorithm using the Tensor Lab package [21]. We generated our sources in the same way as in the original SOBIUM paper - by first generating them as complex Gaussian random variables, and then filtering them using a row of a 16×16 Hadamard matrix - specifically the rows 1,2,4,7, and 8. S is a 12×10000 matrix, of which 5 of the rows are non-zero, each of the non-zero rows are a different one of the filtered sources. A is a 4×12 matrix with entries drawn from a complex Gaussian distribution. Noise was added in the form of additive white Gaussian noise. The results are shown in Figures 3 and 4. Figure 3 shows how the success rate changes as the noise varies, and Figure 4 shows how the success rate varies as the number of samples changes with a SNR of -5 dB.

C. ALESCAF

For testing our approach with ALESCAF, we used the implementation of ALESCAF from the CAF toolbox [22], and used fourth order derivatives only.

Our sources were generated using the CAF toolbox using the 4psk option with a support of $\{-1, 1\}$. S is a 8×256 matrix of which 4 of the rows are non-zero. A is a 2×8 matrix with

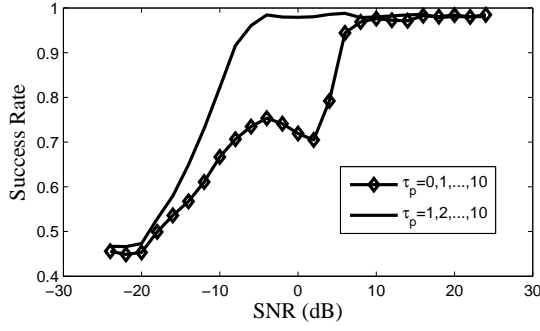


Figure 3. Success Rate of recovering sparsity profile with SOBIUM

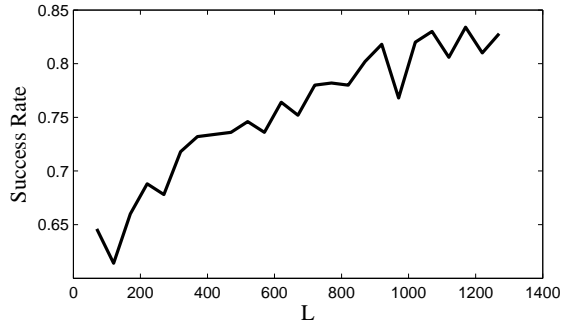


Figure 4. Success rate of recovering sparsity profile with SOBIUM as L varies

entries drawn from a Gaussian distribution. Noise was added in the form of additive white Gaussian noise.

The results can be seen in Figures 5 and 6. Figure 5 shows how the success rate changes as the noise varies, and Figure 6 shows how the success rate varies as the number of samples changes with a SNR of 50 dB.

VI. CONCLUSIONS

In this work, we have developed a general framework for recovering the sparsity profile with a minimal number of

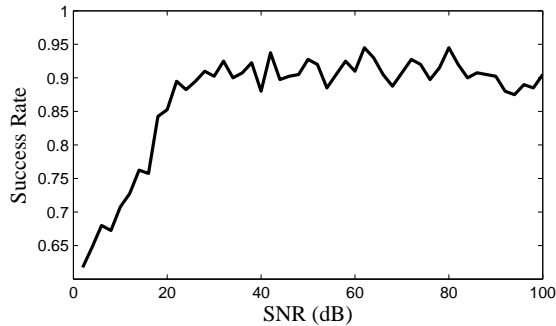


Figure 5. Success Rate of recovering sparsity profile with ALESCAF as noise varies

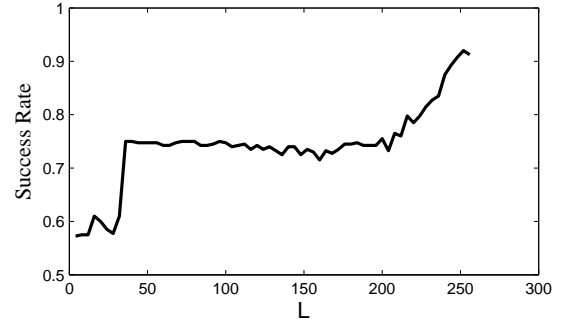


Figure 6. Success rate of recovering sparsity profile with ALESCAF as L varies

sensors in the MMV model by using blind identification methods, and we have demonstrated successful results using a range of data and blind identification methods. In doing so, we have also demonstrated that minimisation of the ℓ_0 pseudonorm is not always the optimal approach.

The results with the FastICA and ALESCAF algorithms provide a strong motivation for investigating non-Gaussian models for compressed sensing, which some researchers have already begun to investigate [23], and the results with the SOBIUM algorithm similarly provide motivation for using compressed sensing models which take into account temporal correlation, as has also been begun to be investigated [24].

Although our framework cannot provide the absolute guarantees that ℓ_0 pseudonorm minimisation can provide, it should be noted that those guarantees are only available in the unrealistic noise free case. Further, ℓ_0 pseudonorm minimisation is a computationally difficult problem, and so approximate methods are used, with ℓ_1 minimisation being one of the most popular techniques used in compressed sensing.

REFERENCES

- [1] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *Information Theory, IEEE Transactions on*, vol. 52, no. 2, pp. 489–509, 2006.
- [2] D. L. Donoho, "Compressed sensing," *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [3] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *Signal Processing, IEEE Transactions on*, vol. 53, no. 7, pp. 2477–2488, 2005.
- [4] J.-P. Roll, "Contribution à la proprioception musculaire, à la perception et au contrôle du mouvement chez l'homme," Ph.D. dissertation, Université d'aix Marseille, 1981.
- [5] B. B. Biswal and J. L. Ulmer, "Blind source separation of multiple signal sources of fMRI data sets using independent component analysis," *Journal of computer assisted tomography*, vol. 23, no. 2, pp. 265–271, 1999.
- [6] D. Boutte and B. Santhanam, "A hybrid ICA-SVM approach to continuous phase modulation recognition," *Signal Processing Letters, IEEE*, vol. 16, no. 5, pp. 402–405, 2009.
- [7] H. Taigang, G. Clifford, and L. Tarassanko, "Application of ICA in removing artefacts from the ECG," *Neural Processing Letters*, vol. 10, no. 1, pp. 1–5, 2004.

- [8] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *Neural Networks, IEEE Transactions on*, vol. 10, no. 3, pp. 626–634, 1999.
- [9] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non-gaussian signals," in *IEE Proceedings F (Radar and Signal Processing)*, vol. 140, no. 6. IET, 1993, pp. 362–370.
- [10] A. M. Kagan, C. R. Rao, and Y. V. Linnik, "Characterization problems in mathematical statistics," *Probability and Mathematical Statistics*, 1973.
- [11] P. Comon and M. Rajih, "Blind identification of under-determined mixtures based on the characteristic function," *Signal Processing*, vol. 86, no. 9, pp. 2271–2281, 2006.
- [12] L. De Lathauwer and J. Castaing, "Second-order blind identification of underdetermined mixtures," in *Independent Component Analysis and Blind Signal Separation*. Springer, 2006, pp. 40–47.
- [13] P. Bofill and M. Zibulevsky, "Underdetermined blind source separation using sparse representations," *Signal processing*, vol. 81, no. 11, pp. 2353–2362, 2001.
- [14] P. Comon, "Blind identification and source separation in 2×3 under-determined mixtures," *Signal Processing, IEEE Transactions on*, vol. 52, no. 1, pp. 11–22, 2004.
- [15] L. De Lathauwer, B. De Moor, and J. Vandewalle, "An algebraic ICA algorithm for 3 sources and 2 sensors," in *Proc. Xth European Signal Processing Conference (EUSIPCO 2000)*, 2000, pp. 5–8.
- [16] M. A. Khajehnejad, W. Xu, A. S. Avestimehr, and B. Hassibi, "Weighted ℓ_1 minimization for sparse recovery with prior information," in *Information Theory. ISIT 2009. IEEE International Symposium on*. IEEE, 2009, pp. 483–487.
- [17] N. Vaswani and W. Lu, "Modified-CS: Modifying compressive sensing for problems with partially known support," *Signal Processing, IEEE Transactions on*, vol. 58, no. 9, pp. 4595–4607, 2010.
- [18] R. Baraniuk, M. A. Davenport, M. F. Duarte, and C. Hegde, "An introduction to compressive sensing," *Connexions e-textbook*, 2011.
- [19] A. Taleb, "An algorithm for the blind identification of N independent signals with 2 sensors," *International Symposium on Signal Processing and its Applications (ISSPA)*, 2001.
- [20] H. Gavert, J. Hurri, J. Sarela, and A. Hyvarinen, "FastICA Matlab package," *Laboratory of Computer and Information Science, Available online*, URL: <http://research.ics.aalto.fi/ica/fastica/t>, 2005.
- [21] L. Sorber, M. Van Barel, and L. De Lathauwer, *Tensorlab v2. 0*, Available online, URL: <http://www.tensorlab.net>, 2014.
- [22] X. Luciani, *CAF toolbox v1.0*, Available online, URL: <http://lucianix.free.fr>, 2015.
- [23] A. Achim, B. Buxton, G. Tzagkarakis, and P. Tsakalides, "Compressive sensing for ultrasound RF echoes using α -stable distributions," in *Engineering in Medicine and Biology Society (EMBC), Annual International Conference of the IEEE*. IEEE, 2010, pp. 4304–4307.
- [24] Z. Zhang and B. D. Rao, "Sparse signal recovery with temporally correlated source vectors using sparse bayesian learning," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 5, no. 5, pp. 912–926, 2011.